

**A Note on Optimal Bayesian  
Design for Nonlinear Problems**

**By**

**Kathryn Chaloner  
School of Statistics, University of Minnesota**

**Technical Report #572**

**February 21, 1992**

# A NOTE ON OPTIMAL BAYESIAN DESIGN FOR NONLINEAR PROBLEMS

Kathryn Chaloner

*School of Statistics, University of Minnesota,  
St. Paul, MN 55108, U.S.A.*

## Abstract

One parameter design problems with a single explanatory variable are considered. Conditions are derived under which a one-point design is optimal in a Bayesian sense under a prior distribution on the parameter. In addition, optimal Bayesian designs for symmetric prior distributions with two support points are characterized for logistic regression with known slope parameter.

*AMS Subject Classification:* Primary 62K05; Secondary 62F15.

*Key Words:* Bayesian design; logistic regression; nonlinear models; optimal design.

## 1 Introduction

Independent responses, not necessarily normally distributed, may be observed at values  $x$  of a single explanatory variable. The response depends on a single parameter, denoted  $\theta$ . The design problem is to choose values  $x_i, i = 1, \dots, n$ , of the explanatory variable. The expected Fisher information for a design putting just one observation at  $x$  will be denoted  $I(\theta, x)$ . For a design putting  $n_i$  observations at  $x_i, i = 1, \dots, k$ , where  $\sum n_i = n$ , the Fisher information is

$$\sum_{i=1}^k n_i I(\theta, x_i) = n \sum_{i=1}^k \frac{n_i}{n} I(\theta, x_i).$$

The usual practice of optimal design theory (eg. Silvey, 1980) is to think of the design as a probability measure  $\eta(x)$ . That is  $\eta(x_i) = n_i/n$ . Define  $H$  to be the set of probability measures,  $\eta(x)$ , over a compact design region  $\mathcal{X}$ . The optimal design

is then the probability measure in  $H$  optimizing some function of the information. For a probability measure  $\eta$  the normalized Fisher information is denoted  $I(\theta, \eta)$  where

$$I(\theta, \eta) = \int I(\theta, x) d\eta(x).$$

Bayesian optimal designs maximize the expectation over a prior distribution of some function of  $I(\theta, \eta)$ , the function being chosen to approximate some utility function. For example Chaloner (1987) and Chaloner and Larntz (1989, 1992) use the two criteria  $\phi_1(\eta) = E[\log I(\theta, \eta)]$  and  $\phi_2(\eta) = -E[I(\theta, \eta)]^{-1}$ , where the expectation is over a prior distribution on  $\theta$ . Under an asymptotic normal approximation to the posterior distribution of  $\theta$ , the criterion  $\phi_1(\cdot)$  approximates the expected Shannon information and the criterion  $\phi_2(\cdot)$  approximates minus the square error loss. Other criterion functions  $\phi(\cdot)$  are also meaningful. Many, like  $\phi_1(\cdot)$  and  $\phi_2(\cdot)$ , are concave on  $H$ .

Another class of criteria are called local optimality criteria and require a best guess of  $\theta$ ,  $\theta_0$  say (Chernoff 1953). The locally optimal design is that which maximizes some function of the information evaluated at  $\theta_0$ . For example,  $\phi_1^\theta(\eta) = \log I(\theta_0, \eta)$  and  $\phi_2^\theta(\eta) = -[I(\theta_0, \eta)]^{-1}$  are local optimality criteria analagous to the Bayesian criteria  $\phi_1(\cdot)$  and  $\phi_2(\cdot)$ . Locally optimal designs can be thought of as Bayesian designs for one point prior distributions. Local optimality criteria will be denoted as  $\phi^\theta(\eta)$  to denote the dependence of the criterion on a single value of  $\theta$ . A Bayesian criterion can be expressed as  $\phi(\eta) = E\phi^\theta(\eta)$ .

Chaloner and Larntz (1989) give an equivalence theorem and show how it can be used to verify that a design is optimal for such criteria. They show how a result of Whittle (1973) can be generalized to apply to nonlinear design problems. They also show empirically how, in several examples, as the uncertainty in the prior distribution increases so does the number of support points of the optimal design measure. The equivalence theorem is given in the appendix and will be used in Section 3.

There are many examples of locally optimal designs derived in closed form as functions of  $\theta_0$ : for example in Kitsos, Titterton and Tornsey (1988). For non-degenerate prior distributions, however, the only examples of Bayesian optimal designs with more than one support point are derived numerically and cannot be expressed in closed form. For practical design problems it is reasonable to rely on numerical programs to find Bayesian designs but for understanding the Bayesian criteria it is useful to study some simple examples and closed form results.

In Section 2 sufficient conditions are given for a one point design to be optimal both under local optimality and under a Bayesian criterion with a non-degenerate prior distribution. The conditions require that the support of the prior distribution is small. This result has practical implications because if the locally optimal design is optimal for a non-degenerate prior distribution this is a good property of the locally optimal design. It implies that if there is uncertainty in the best guess  $\theta_0$  the locally optimal design may still be optimal in a Bayesian sense. The result is applied to a nonlinear regression model.

In Section 3  $\phi_1$ -optimal designs for logistic regression with a known slope are studied. Closed form expressions are given for  $\phi_1$ -optimal designs for a class of prior distributions. The prior distributions have exactly two support points with equal mass. It is demonstrated exactly how, for this simple case, the optimal design changes from a measure with one support point to one with two support points.

The prior distribution in this example has only two points of support which is clearly not very realistic. In particular, if the prior distribution really did have only two support points then the posterior distribution would also be concentrated on the same two points; the asymptotic normality assumption required to justify the criterion would be inappropriate. This example does, however, provide closed form solutions that should lead to better understanding of more realistic problems in which the prior distribution is continuous. In addition, the example can be justified by thinking of the criterion as applying when a continuous vague prior

distribution will be used for inference but an informative prior distribution will be used for design. This is motivated in Tsutakawa (1972).

Bayesian design for nonlinear models is much harder than in linear models as exact expressions for expected utilities are difficult to calculate. Bayesian design for linear models is discussed in Pilz (1983), Chaloner (1984), DasGupta and Studden (1991) El-Krunz and Studden (1991) and elsewhere.

## 2 Conditions for one point designs

### Theorem 1

Consider a one parameter design problem with a prior distribution with support  $\Theta$  and a design region  $\mathcal{X}$ . The set  $H$  is the set of probability measures on  $\mathcal{X}$ . Suppose that the following conditions hold:

1. For each  $\theta$  in  $\Theta$  the function  $I(\theta, x)$  is concave in  $x$  over an interval  $[a(\theta), b(\theta)]$  and strictly concave over  $(a(\theta), b(\theta))$  with a unique value  $x = m(\theta)$  maximizing  $I(\theta, x)$  where  $m(\theta)$  is in  $[a(\theta), b(\theta)]$ . Furthermore, for each  $\theta$  in  $\Theta$ , if  $x < a(\theta)$  then  $I(\theta, a) > I(\theta, x)$  and if  $x > b$  then  $I(\theta, b(\theta)) > I(\theta, x)$ .
2. Define  $a = \sup_{\theta} a(\theta)$ ,  $b = \inf_{\theta} b(\theta)$ ,  $c = \inf_{\theta} m(\theta)$  and  $d = \sup_{\theta} m(\theta)$ . The support of the prior distribution is such that  $a < c \leq d < b$ .
3. The interval  $[c, d]$  is a subset of  $\mathcal{X}$ .

Then, for any Bayesian optimality criterion  $\phi(\cdot)$  which is concave on  $H$ , there is a unique optimal one point design with support in  $[c, d]$ .

Furthermore, the corresponding locally optimal design for any best guess  $\theta_0$  is a one point design putting mass one at  $m(\theta_0)$ .

### Proof

Consider an  $I(\theta, x)$  satisfying condition 1 and a prior distribution with support satisfying condition 2. It will first be shown that there is an optimal Bayesian

design with support in the interval  $[c, d]$ . Take any design  $\eta_1$  which puts mass outside of the interval  $[c, d]$ . Construct a design  $\eta_2$  which is identical to  $\eta_1$  except that the mass at  $x < c$  is moved to  $c$  and the mass at  $x > d$  is moved to  $d$ ; then clearly  $\phi(\eta_2) > \phi(\eta_1)$ . An optimal Bayesian design measure must therefore have all its support in  $[c, d]$ .

Now consider any design  $\eta_2(x)$  with support in  $[c, d]$ . The function  $I(\theta, x)$  is strictly concave in  $x$  over  $[c, d]$  for each  $\theta$ . Construct a one point design  $\eta_3(x)$  which has unit mass at the point  $\bar{x}$  where  $\bar{x} = \int x d\eta_2(x)$ . By concavity of  $I(\theta, x)$   $\phi^\theta(\eta_3) \geq \phi^\theta(\eta_2)$  for each  $\theta$  and hence  $\phi(\eta_3) \geq \phi(\eta_2)$ ; the inequality is strict unless  $\eta_2(x)$  has one support point. Thus for any design with more than one support point there is a design with one support point which is better. If condition 3 is satisfied there must, therefore, be a unique optimal Bayesian design with one support point.

Under local optimality with a best guess  $\theta_0$  it is easy to see that the locally optimal design takes all observations at  $x = m(\theta_0)$ . This completes the proof.

Note that the first condition is such that  $I(\theta, x)$  is essentially unimodal. The second condition requires that  $\Theta$  is small. These conditions can be easily satisfied for several problems, including binary regression, and the nonlinear regression example to follow.

## 2.1 Example

Consider a non-linear regression model where independent observations  $y_i$  are such that

$$y_i = e^{-\theta x} + e_i$$

and the  $e_i$  are independent, normally distributed and have mean zero and variance one. The explanatory variable  $x$  is constrained to be positive. The information function is  $I(\theta, x) = x^2 e^{-2\theta x}$  and any local optimality criterion gives a one-point optimal design with all mass at the point  $x = \theta_0^{-1}$  where  $\theta_0$  is the best guess

(Pronzato and Walter, 1985). It can also be shown that

$$\begin{aligned} a(\theta) &= \frac{2 - \sqrt{2}}{2\theta} \\ b(\theta) &= \frac{2 + \sqrt{2}}{2\theta} \\ m(\theta) &= \frac{1}{\theta} \end{aligned}$$

Furthermore,  $a(\theta) < m(\theta) < b(\theta)$ . Condition 2 of Theorem 1 is satisfied if and only if

$$\frac{\sup_{\Theta} \theta}{\inf_{\Theta} \theta} \leq \frac{2 + \sqrt{2}}{2}. \quad (1)$$

So if  $\Theta$  is small enough for (1) to be satisfied the optimal design for any Bayesian criterion is a one point design. For  $\phi_1$ -optimality it is also easy to show that such an optimal one point design puts all mass at  $x = [E(\theta)]^{-1}$ .

### 3 $\phi_1$ -optimality for logistic regression

Suppose that responses at a value  $x$  of the explanatory variable are independent Bernoulli random variables with probability of response  $p(x)$  where

$$p(x) = \frac{1}{[1 + \exp\{-\beta(x - \theta)\}]}$$

The parameter  $\theta$  is the value of  $x$  where  $p(x) = \frac{1}{2}$ , and is also known as the LD50, and  $\beta$  is the slope in the logit scale. The value of  $\theta$  will be taken to be unknown and the slope parameter  $\beta$  known. Without loss of generality let  $\beta$  be equal to one.

The normalized Fisher information for a design measure  $\eta$  is

$$I(\theta, \eta) = \sum \eta(x_i) I(\theta, x_i) = \sum \eta(x_i) \frac{e^{x_i - \theta}}{(1 + e^{x_i - \theta})^2}.$$

$I(\theta, x)$ , as a function of  $x$ , is symmetric around  $\theta$ , and is concave in  $x$  in the interval  $\theta \pm \log(2 + \sqrt{3})$ .

The locally optimal design, with  $\theta = \theta_0$ , is clearly to put all observations where  $I(\theta_0, x)$  is maximized: that is take all observations at  $x = \theta_0$ . A Bayesian  $\phi_1$ -optimal design will be the design measure  $\eta$ , in  $H$ , maximizing  $E[\log I(\theta, \eta)]$ .

The following theorem characterizes all  $\phi_1$ -optimal designs for prior distributions which put probability  $\frac{1}{2}$  on each of two points. The proof requires the use of the equivalence theorem from the appendix. Without loss of generality the two points in the prior distribution are defined to be  $\pm g$ .

### Theorem 2

For a prior distribution which puts probability  $\frac{1}{2}$  on each of  $\pm g$  the  $\phi_1$ -optimal design is:

1. if  $|g| \leq \frac{1}{2} \log(2 + \sqrt{3})$  the  $\phi_1$ -optimal design measure puts mass 1 at 0.
2. if  $|g| > \frac{1}{2} \log(2 + \sqrt{3})$  the  $\phi_1$ -optimal design measure puts mass  $\frac{1}{2}$  at each of the points  $\pm B(g)$  where

$$B(g) = \log \frac{e^{4g} - 6e^{2g} + 1 + (e^{2g} - 1)\sqrt{e^{4g} - 14e^{2g} + 1}}{2(e^{3g} + e^g)}. \quad (2)$$

### Proof

The proof is straightforward using the derivative function to verify optimality. For  $\phi_1$ -optimality the derivative at a design  $\eta$  in the direction  $x$  is

$$d(\eta, x) = E \frac{I(\theta, x)}{I(\theta, \eta)} - 1.$$

For the two point design  $\eta$  which puts mass  $\frac{1}{2}$  at each of  $\pm B(g)$  given in (2), the derivative, for the two point prior simplifies to:

$$-\frac{(e^{2x+2g} + e^{2x+g} - e^{x+4g} + 6e^{x+2g} - e^x + e^{3g} + e^g)^2}{(e^{2g} + 1)^2(e^x + e^g)^2(e^{x+g} + 1)^2}$$

This is clearly always non-positive and so the design putting mass  $\frac{1}{2}$  at  $\pm B$  must be optimal. Note also that the expression under the square root sign in (2) is strictly positive only if  $|g| > \frac{1}{2} \log(2 + \sqrt{3})$ , in which case the derivative has two roots, at  $\pm B(g)$ .



For  $|g| \leq \frac{1}{2} \log(2 + \sqrt{3})$  the result of Theorem 1 could be used, together with symmetry, to prove that the optimal design is point mass at zero. Alternatively, the design which puts all mass at zero has derivative:

$$-\frac{(e^x - 1)^2(2e^{2x+2g} - e^{x+4g} + 2e^{x+3g} + 2e^{x+2g} + 2e^{x+g} - e^x + 2e^{2g})}{2(e^x + e^g)^2(e^{x+g} + 1)^2}.$$

The term in the numerator is  $(e^x - 1)^2$  multiplied by a quadratic in  $e^x$ . It is straightforward to show that, if  $|g| \leq \frac{1}{2} \log(2 + \sqrt{3})$ , the quadratic is positive and the derivative has a single root at  $x = 0$ . The result is proved.

Note that if  $|g| = \frac{1}{2} \log(2 + \sqrt{3})$  Equation (2) gives  $B(g) = 0$ . If  $|g|$  converges to  $\frac{1}{2} \log(2 + \sqrt{3})$  from above then  $|B(g)|$  converges to 0 from above and the change from an optimal design on two support points to one on one support point therefore occurs in a continuous manner.

The result of Theorem 1 can be used to extend this result so that any prior distribution with support contained in an interval of length less than  $\log(2 + \sqrt{3})$  gives a one point optimal design.

## Acknowledgement

This work was supported by a grant from the CNR at the University of Perugia, Italy. I am especially grateful to Alessandra Giovagnoli for very helpful comments.

## Appendix

The criterion  $\phi(\eta)$  is a concave function over the set  $H$  of probability measures. Define  $D(\eta_1, \eta_2)$  to be the derivative of  $\phi(\eta)$  at  $\eta_1$  in the direction of  $\eta_2$ . That is

$$D(\eta_1, \eta_2) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} [\phi\{(1 - \epsilon)\eta_1 - \epsilon\eta_2\} - \phi(\eta_1)]$$

Define  $\eta_x$  to be the probability measure which is point mass at  $x$  in  $X$ , then the function  $D(\eta, x)$  is denoted to be  $d(\eta, \eta_x)$ .

The following equivalence theorem gives conditions for a design  $\eta_0$  to be optimal.

### Equivalence Theorem

a) An optimal design  $\eta_0$  can be equivalently characterized by any of the three conditions

1.  $\eta_0$  minimizes  $\phi(\cdot)$
2.  $\eta_0$  minimizes  $\sup_{x \in X} d(\eta, x)$
3.  $\sup_{x \in X} d(\eta, x) = 0$ .

b) The point  $(\eta_0, \eta_0)$  is a saddlepoint of  $D$  in that  $D(\eta_0, \eta_1) \leq 0 = D(\eta_0, \eta_0) \leq D(\eta_2, \eta_0)$  for all  $\eta_1, \eta_2$ .

The theorem can be applied to any concave criterion and the proof is outlined in Whittle (1973) for linear models; Chaloner and Larntz (1989) show how it can be applied to nonlinear Bayesian design criteria.

## REFERENCES

- Chaloner, K. (1984). Optimal Bayesian experimental design for linear models. *Ann. Statist.* 7, 283-300.
- Chaloner, K. (1987). An approach to design for generalized linear models. In *Model Oriented Data Analysis*, V.V. Fedorov and H. Läuter eds.. Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, 3-12.
- Chaloner, K. and Larntz, K. (1989). Optimal Bayesian experimental design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21, 191-208.
- Chaloner, K. and Larntz, K. (1992). Bayesian design for accelerated life testing. To appear in the *Journal of Statistical Planning and Inference*.
- Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* 24, 586-602.

- DasGupta, A. and Studden, W.J. (1991). Robust Bayesian experimental designs in normal linear models. *Ann. Statist* 19, 1244-1256.
- El-Krunz, S.M. and Studden, W.J. (1991). Bayesian optimal designs for linear regression models. *Ann. Statist.* 19, 2183-2208.
- Kitsos, C.P., Titterington, D.M. and Tornsey, B. (1988). An optimal design problem in Rhythmometry. *Biometrics*, 44, 657-671.
- Pilz, J. (1983). *Bayesian estimation and experimental designs in linear regression models*. Teubner-Texte, Leipzig.
- Pronzato, L. and Walter, E. (1985). Robust experiment design via stochastic approximation. *Mathematical Biosciences*, 75, 103-120.
- Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall, London.
- Tsutakawa, R.K. (1972). Design of an experiment for bioassay. *J. Amer. Statist. Assoc.* 67, 584-590.
- Whittle, P. (1973). Some general points in the theory and construction of D-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* 35, 123-130.